RESEARCH ARTICLE OPEN ACCESS

A Survey on Location Aware Keyword Query Suggestion Based On Document Proximity

*Twinkle Pardeshi, Prof. Pradnya Kulkarni
Department of Computer Engineering Maharashtra Institute of Technology, Pune, Maharashtra, India
Corresponding Author: Twinkle Pardeshi

ABSTRACT
Keyword suggestion in web search is a very important feature to be considered in today's growing world. It helps user to access the information without any priorknowledge of how to express in queries.The main concept of query suggestion is used to retrieve documents from the related server by consuming less time. Platform is provided by search engines for users to describe their information need more precisely by using query recommendation. Previously there has been lot of work done for retrieving relevant data of users to meet their information need and improvingperformance of search engines. This paper reviewsand compares different available methods in querylogprocessing for information retrieval.Thenconclude that Existing keyword suggestion techniques are not considering the locations of theusers and the query results which serves as a drawback of existingsystems. The spatial factor is not consideredto retrieve result. The approach based on location aware keyword query suggestion is better to understand user'sinteraction process with search engines to find theappropriate information need.

Keywords: Data mining, Spatial location, keyword suggestion.

Date of Submission: 13-08-2017 Date of acceptance: 09-09-2017

I. INTRODUCTION

Keyword suggestion has become a key feature of commercial Web search engines. Many a times after submitting a keyword query, the user may not be satisfied with the results as the expected results were not retrieved. Major problem of current search engine is that the queries are short. User submits short queries and expect relevant results prior that they have very less knowledge about the query. Most of the times users are not capable of expressing queries precisely but hunt of the relevant results. Search engines must be intelligent to understand what user wants. The list of keywords displayed by the web search engine is not always good descriptor of the information needs of user.

To overcome the problem, many search engines have implemented the query suggestion method, Also known as keyword suggestion. Effective keywordsuggestion methods are based on click information from query logs, and query session data or query topic models. None of the techniques are location based i.e. none of them provide location based relevant retrieval of information. Users wants that the search engine must able to retrieve the information which is relevant and also near to the location of user. The keyword suggestion helps user to search or to access the relevant information. The keyword suggestion module of the search engine recommends a set of keywords queries that are most likely to refine the users search. Effective method for keyword suggestion is based on information from query log.

With billions of pages on the Web, trying to find the right words to use when youwant to search for something can often be hard, especially when you are looking for information on a topic that you don't know too much about. Search engines often act asa middle man between searcher and site owner, helping to bring people to pages that may help them satisfy some kind of informational need, or accomplish some task.Search engines will often offer search suggestions based upon the words that a searcher types into a search box, to try to make it easier for those searchers. Knowing more about how search engines find those suggestions may be helpful to searchers and to site owners. When user clicks the url out of the search results list, the information about user query and the clicked url are stored in the server log. Get the related information from search engine logs. All the information forms a working set. Learn aspects from the information in the working set. These aspects correspond to usersinterests given the input query. Each aspect is labeled with a representative query. Categorize and organize the search results of the input query according to the aspects learned above. When next time user enters the query, the results are retrieved from search engines and compared with the data saved in the

server log and rank the search results accordingly so that users can reach effortlessly what they are looking for. However, even though different search intents behind a given query may have been popular at different time periods in the past, existing query suggestion methods neither utilize nor present such information.

II. LITERATURE SURVEY

Data mining, the extraction of hidden predictive information from large databases,is a powerful new technology with great potential to help companies focus on the most important information in their data warehouses. Data mining tools predict future trends and behaviors, allowing businesses to make proactive, knowledge-driven decisions. Data mining allows you to sift through all the chaotic and repetitive noise in your data, understand what is relevant and then make good use of that information to assess likely outcomes, accelerate the pace of making informed decisions.

The location aware keyword(LKS) query suggestion method provide the suggested queries retrieve documents which is related to user information and located near to users location.LKS framework, it construct and use keyword document bipartite graph (KD graph) that connect to keyword queries with their relevant document. LKS adjust weight on edges in KD graph to capture the semantics relevance between keyword queries and spatial distance between document location and user location. For distance calculation the Personalized PageRank(PPR) algorithm is used, it uses Random walk with restart(RWR) on KD graph, starting from user supplied query to find the set of keywords and spatial proximity to the user location. Keyword suggestion is used to return relevant document to the user. Sometimes, user of the system are not satisfied with the results of search engines as they fail to provide useful information to user. Location-based document retrieval is introduced by Shuyao qi, Dingming wu, and Nikos Mamoulis which aims at retrieving documents not only based on relevance but also location aware documents. Location-aware Keyword query Suggestion framework is introduced for suggestions relevant to the users information needs that also retrieve relevant documents close to the query user's location. The state-of-the-art Bookmark Coloring Algorithm for RWR search is extended to compute the location-aware suggestions. In addition, a partition-based algorithm is being proposed that greatly reduces the computational cost of BCA. bookmark-coloring algorithm (BCA) computes weights over the web pages utilizing the web hyperlink structure. For page-specific PageRank computing, bookmark-coloring model is considered. BCV can be considered as an efficient and sparse version of a page-specific PageRankthat leads to an almost identical ordering of search results. BCA utilizes local propagation it never touches certain distant nodes. So, BCA is muchmore focused than PageRank. P. Berkhin ,states that BCA also has important algebraic properties and if several BCVs corresponding to a set of pages (called hub) are known, they can be leveraged in computing arbitrary BCV via a straightforward algebraic process and hub BCVs can be efficiently computed and encoded. Personalized PageRank (PPR) , also known as Random Walk with Restart (RWR), has emerged as the most popular link-based similarity measure due to its effectiveness and solid theoretical foundation. It is important to set PPR parameters in an ad-hoc manner when finding similar nodes because of dynamically changing nature of graphs. Through interactive actions, interactive similarity search supports users to enhance the efficiencyof applications. Unfortunately, if the graph is large, interactive similarity search is infeasible due to its high computation cost. Hence, to deal with this problem solution is Castanet. It is important to iteratively find the top-k nodes for an arbitrary graph in an ad-hoc manner. In this approach, users are allowed to adjust the scaling parameter for a given graph on the fly. One of the most successful techniques known to the academic communities is based on random walk with restart. This is because the proximity defined by RWR yields the following benefits such as it captures the global structure of the graph and captures multi-facet relationships between two nodes unlike traditional graph distances. The computation of the proximities by RWR is computationally expensive. Random walk with restart provides a good proximity score between two nodes in a graph, and it hasbeen successfully used in many applications such as automatic image captioning, recommender systems, and link prediction. The goal is to find nodes that have top k highest proximities for a given node. Solution to the problem is K- dash, that computes the proximity of a selected node efficiently by sparse matrices and it skips unnecessary proximity computations when searching for the top-k nodes.

User is required to type full query which become irrelevant for user to do so. The method is introduced in the paper which does not demand for query completions and to select the desired query, no clicking is required to see the final results here. A Type Ahead Search System is proposed in this paper. It has to be responsive and must look instantaneous to the user. Integrated architecture for location-aware TAS is proposed. Furthermore, it suggests novel techniques to augment that architecture with materialized information for efficient queryprocessing.The main challenge in Location aware instant search, is to achieve high interactive speed. Users are required to enter full query keyword which many a times becomes irrelevant for users . To

address this challenge,, a novel index structure is proposed , prefix region tree (called PR-Tree), to efficiently support location aware instant search. PR-Tree is a tree-based index structure which seamlessly integrates the textual description and spatial information to index the spatial data. Using the PR-Tree, efficient algorithms have being developed to support single prefix queries and multi-keyword queries. Prefix-regions is used to partition the spatial data by considering the spatial information and textual description simultaneously. An intersection-basedalgorithm and a cost-based algorithm has being developed to answer multi-keyword queries.Spatial keyword queries consider both locations and textual descriptions of the objects. It is truly important to know there needs. Relevance of an object to a query is measured by spatial-textual similarity that is based on both spatial proximity and textual similarity. Y. Lu, J. Lu, G. Cong introducedthe Reverse Spatial-Keyword k-Nearest Neighbor query, which finds those objects that have the query as one of their k-nearest spatial-textual objects. To deal with Reverse Spatial-Keyword k-Nearest Neighbor queries, a hybrid index tree is proposed, called IUR-tree (Intersection- Union R-tree) that effectively combines location proximity with textual similarity. To accelerate the query processing, IUR-tree is improved by leveraging thedistribution of textual description, leading to some variants of the IUR-tree called Clustered IUR-tree (CIUR-tree) and combined clustered IqUR-tree (C2IUR-tree), for each of which optimized algorithms are developed.

Motivation

Web mining is the application of data mining techniques to extract knowledge from web data, including web documents, hyperlinks between documents, etc. Mining in such a hug data is very difficult. User issuing query is in need of information related to his query and providing user with data of his interest is very important task. Users needs are to get the relevant documents related to its search query so that less amount of time is invested for searching. Nowadays, query suggestion is the important factor that cannot be ignored. End user inputs the query, and expect the intelligent search engines to suggest relevant results as per end users need. Sometimes end user fails to express relevant query which may lead to unsatisfactory results. Hence, there comes a need to develop a model that consider such problems. Even if the results are relevant they are not nearer to end users location. Location based retrieval is as important as relevant information. There are various existing methods that do not consider user's location while retrieving the query results. There is need to not only retrieve documents related to the user information needs but also located near the user location.

Architecture

A. Keywork Query Sugesstion

Query suggestion does not take into account the location of the query issuer which leads to unsatisfactory results. LSK i.e. location-aware keyword query suggestion is introduced such that the suggested queries retrieve results considering both user informational needs and located nearby the users location. There are basically two challenges of LSK framework, those are:

a) To effectively measure keyword query similarity while capturing the spatial distance factor.
b) To compute the suggestions efficiently.

Approaches for keyword query suggestion are classified into 3 categories, those are:

a) Random walk based approaches
   This method uses graph structure for modeling the information that is provided by query log and then applies the random walk process on graph for query suggestion.
b) Cluster based approaches
   In this method the query log is viewed as query URL bipartite graph. By applying the clustering algorithm on vertices in the graph, query cluster can be identified. Then, user supplied query q and queries that are belonging to same cluster as q does not returned to the user as suggestion
c) Learning to Rank Approaches
   This approach is trained based on different type of query features like query performance prediction. Given query q, a list of suggestion is produced based on their similarity to q in topic distribution space.

I. Keyword-Document Graph

Location aware keyword query suggestion constructs keyword document graph. The graph G(D,K,E) is the directed weighted bipartite graph between document(D) and keyword(K) that captures the semantics similarity and textual relevance between the keyword query and document nodes; i.e., the satisfies the first criterion of location- aware suggestion. In the graph E contains edge from particular keyword to document and another edge from that particular document to keyword node. Random walk with restart is applied on KD graph to find set of m keyword queries with high semantic relevance and spatial proximity to users location. In most of the applications working with RWR search require transition probabilities between nodes beforehand. However, the edge weights of our KD-graph are unknown in advance. RWR search have high cost for large graphs, hence partition-based algorithm is introduced that reduces the cost of RWR search on dynamic bipartite graph.

II. Location Aware Edge Weight Adjustment

The KD graph is constructed to provide location based retrieval. Edge adjustment is done according to query which is dynamic in nature. Adjustment to one edge does not affect another. Main motive of adjustment is to locate document at minimum distance from the query issuer.

III. Location Aware Keyword Query Suggestion

KD graph is constructed only once. Hence while adjusting the edges a new graph is created from original graph. Once the adjustment of edges between keyword and document is done, next question comes in that how the suggestion will be returned based on the issuers location. Query suggestions must satisfy two criteria i.e. relevance and closeness to query. Proximity score is computed with respect to query keyword based on random walk with restart process. RWR score of node v in newly created graph after adjustment models probability that random surfer starting from keyword query will reach node v. At every step, the surfer will move either to the adjacent node or teleports to query keyword node. The top-m nodes with the highest score are the suggestions.

III. CONCLUSION

Due to the emerging use of search engine it is very important to provide best results to the end user of the search engine. There are number of techniques that are currently being used to retrieve relevant documents according to query issuers interest. When issuer issues a query, crawler of the engine crawls to fetch relevant documents according to the query and return it back to the user or query issuer. After submitting a keyword query, the user may not be satisfied with the results, so the keyword suggestion module of the search engine recommends a set of m keyword queries that are most likely to refine the users search in the right direction. Many times query issuer requires the search results nearby their current location. This requirement is of location based document retrieval is not made available in existing techniques. Hence, Location based Keyword search is studied so that query issuer is provided with location based results. Hence, we have presented a Location-aware keyword query suggestion frame work that provides relevant keyword suggestion to the user and at the same time can retrieve location based documents which improves the quality of keywordsuggestion.

REFERENCE

[1]. R. Baeza-Yates, C. Hurtado, and M. Mendoza, Query recommendation using query logs in search engines,in Proc. Int. Conf. Current Trends Database Technol., 2004, pp. 588-596.

[2]. R. Baeza-Yates, C. Hurtado, and M. Mendoza, Query recommendation using query logs in search engines,in Proc. Int. Conf. Current Trends Database Technol., 2004, pp. 588-596.

[3]. H. Cao, D. Jiang, J. Pei, Q. He, Z. Liao, E. Chen, and H. Li, Context-aware query suggestion by mining click-through and session data, in Proc. 14th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2008, pp. 875-883.

[4]. Q. Mei, D. Zhou, and K. Church,Query suggestion using hitting time, in Proc. 17th ACM Conf. Inf. Knowl. Manage., 2008, pp. 469-478.

[5]. P. Boldi, F. Bonchi, C. Castillo, D. Donato, A. Gionis, and S. Vigna, The query-flow graph: Model and applications, in Proc. 17th ACM Conf. Inf. Knowl. Man- age., 2008, pp. 609-618.

[6]. Y. Lu, J. Lu, G. Cong, W.Wu, and C. Shahabi, Efficient algorithms and cost models for reverse spatial-keyword k-nearest neighbor search, ACM Trans. Database Syst., vol. 39, no. 2, pp. 13:113:46, 2014.

[7]. S. Basu Roy and K. Chakrabarti, Location aware type ahead search on spatial databases: Semantics and efficiency, in Proc. ACM SIGMOD Int. Conf. Manage. Data, 2011, pp. 361-372.

[8]. R. Zhong, J. Fan, G. Li, K.-L. Tan, and L. Zhou, Location-aware instant search, in Proc. 21st ACM Conf. Inf. Knowl. Manage., 2012, pp. 385-394.

[9]. Miliou and A. Vlachou, Location-aware tag recommendations for ickr, in Proc. 25th Int. Conf. Database Expert Syst. Appl., 2014, pp. 97-104.

[10]. H. Tong, C. Faloutsos, and J.-Y. Pan, Fast random walk with restart and its applications, in Proc. 6th Int. Conf. Data Mining, 2006, pp. 613-622.

[11]. Y. Fujiwara, M. Nakatsuji, M. Onizuka, and M. Kitsuregawa, Fast and exact top-k search for random walk with restart, Proc. VLDB Endowment, vol. 5, no.5, pp. 442-453, Jan. 2012.

[12]. B. Bahmani, A. Chowdhury, and A. Goel, Fast incremental and personalized PageRank, Proc. VLDB Endowment, vol. 4, no. 3, pp. 173-184, Dec. 2010.

[13]. P. Berkhin, Bookmark-coloring algorithm for personalized pagerank computing, Internet Math., vol. 3, pp. 41-62, 2006.

[14]. M. Gupta, A. Pathak, and S. Chakrabarti, Fast algorithms for top-k personalized

pagerank queries, in Proc. 17th Int. Conf World Wide Web, 2008, pp. 1225-1226.
[15]. I. S. Dhillon, Co-clustering documents and words using bipartite spectral graph partitioning, in Proc. ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2001, pp. 269-274.
[16]. S. Bhatia, D. Majumdar, and P. Mitra, Query suggestions in the absence of query logs, in Proc. Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval, 2011,pp. 795-804.
[17]. J.-R. Wen, J.-Y. Nie, and H.-J. Zhang, Clustering user queries of a search engine,in Proc. 10th Int. Conf. World Wide Web, 2001, pp. 162-168.
[18]. J. Myllymaki, D. Singleton, A. Cutter, M. Lewis, and S. Eblen, Location based query suggestion, U.S. Patent 8 301 639, Oct. 30, 2012.
[19]. S. Lee, S. Park, M. Kahng, and S.-G. Lee, PathRank: A novel node ranking measure on a heterogeneous graph for recommender systems, in Proc. 21st ACM Int. Conf. Inf. Knowl. Manage., 2012, pp. 1637-1641.
[20]. N. Craswell and M. Szummer, Random walks on the click graph, in Proc. 30th Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval, 2007, pp. 239-246.
[21]. T. Miyanishi and T. Sakai, Time-aware structured query suggestion, in Proc. 36th Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval, 2013, pp. 809-812.

International Journal of Engineering Research and Applications (IJERA) is **UGC approved** Journal with Sl. No. 4525, Journal no. 47088. Indexed in Cross Ref, Index Copernicus (ICV 80.82), NASA, Ads, Researcher Id Thomson Reuters, DOAJ.

Twinkle Pardeshi. "A Survey on Location Aware Keyword Query Suggestion Based On Document Proximity." International Journal of Engineering Research and Applications (IJERA) , vol. 7, no. 9, 2017, pp. 17–21.